Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/EP05/050888

International filing date: 01 March 2005 (01.03.2005)

Document type: Certified copy of priority document

Document details: Country/Office: US
Number: 60/549,219
Filing date: 02 March 2004 (02.03.2004)

Date of receipt at the International Bureau: 12 April 2005 (12.04.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

EP05/50888

PA 1285820

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

February 17, 2005

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE UNDER 35 USC 111.

APPLICATION NUMBER: *60/549,219*
FILING DATE: *March 02, 2004*

By Authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS

N. WILLIAMS
Certifying Officer

PTO/SB/16 (05-03)
Approved for use through 4/30/2003. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53(c).

Express Mail Label No. EF195556712US

INVENTOR(S)

Given Name (first and middle [if any])	Family Name or Surname	Residence (City and either State or Foreign Country)
Lee	Bacheler	Newark, DE 19711, USA

Additional inventors are being named on the 1 *separately numbered sheets attached hereto*

TITLE OF THE INVENTION (500 characters max)

Estimation of clinical cut-offs

Direct all correspondence to: **CORRESPONDENCE ADDRESS**

[X] Customer Number 000027777 → Place Customer Number Bar Code Label here

OR *Type Customer Number here*

[ ] Firm or Individual Name					
Address					
Address					
City		State		ZIP	
Country		Telephone		Fax	

ENCLOSED APPLICATION PARTS *(check all that apply)*

[X] Specification *Number of Pages* 29		[ ] CD(s), Number	
[X] Drawing(s) *Number of Sheets* 8		[ ] Other (specify)	
[ ] Application Data Sheet. See 37 CFR 1.76			

METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT

[ ] Applicant claims small entity status. See 37 CFR 1.27.

[ ] A check or money order is enclosed to cover the filing fees.

[X] The Director is hereby authorized to charge filing fees or credit any overpayment to Deposit Account Number: 10-0750/VIP23/JJ:T

[ ] Payment by credit card. Form PTO-2038 is attached.

FILING FEE AMOUNT ($): $160.00

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

[X] No.

[ ] Yes, the name of the U.S. Government agency and the Government contract number are: _____

Respectfully submitted

SIGNATURE _____

TYPED or PRINTED NAME Jesus Juanos i Timoneda

TELEPHONE (732) 524-1513

Date 03/02/2004

REGISTRATION NO. *(If appropriate)* 43,332

Docket Number: VIP0023

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

PROVISIONAL APPLICATION COVER SHEET
Additional Page

PTO/SB/16 (05-03)
Approved for use through 4/30/2003. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

Docket Number VIP 0023

INVENTOR(S)/APPLICANT(S)		
Given Name (first and middle [if any])	Family or Surname	Residence (City and either State or Foreign Country)
Bart	Winters	B-3650 Dilsen-Stokkem, Belgium

Number 2 of 2

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

DOCKET NO. VIP0023

IN THE UNITED STATES
PATENT AND TRADEMARK OFFICE

Applicant: Lee BACHELER, et al.

For : ESTIMATION OF CLINICAL CUT-OFFS

Express Mail Certificate

"Express Mail" mailing number: EF195556712US

Date of Deposit: March 2, 2004

I hereby certify that this complete provisional application, including specification pages, claims, and drawings, is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to the Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

A Combined Declaration and Power of Attorney will be submitted to the United States Patent and Trademark Office upon receipt of the U.S. Serial Number for this patent application.

Laurie A. Phillips
(Typed or printed name of person mailing paper or fee)

Laurie A. Phillips
(Signature of person mailing paper or fee)

Estimation of clinical cut-offs

The present invention concerns methods and systems for improving the accuracy of predicting resistance of a disease to a drug. More specifically, the invention provides methods for assessing the impact of pre-existing variations in drug susceptibility, whether naturally occurring or selected by previous drug exposure, on treatment response in order to establish clinically relevant cut-off values for phenotypic or genotypic resistance tests.

All publications, patents and patent applications cited herein are incorporated in full by reference.

Techniques to determine the resistance of a disease to a drug are becoming increasingly important. Since the issuance of the first report suggesting a correlation between the emergence of viral resistance and clinical progression, techniques to determine the resistance of a pathogen to a drug have been increasingly incorporated into clinical studies of therapeutic regimens (see Brendan Larder *et al.*, HIV Resistance and Implications for Therapy (1998), herein incorporated by reference). For example, as with viral infections, some studies show that p53 mutations may also be predictive of tumour response to specific anticancer drug therapy, radiation treatment or gene therapy. This is the case in breast cancer where initial studies have shown that cisplatin and tamoxifen are more effective in patients whose tumours have a p53 mutation. Thus, the aim of resistance monitoring is to provide the necessary information to enable the physician to prescribe the most optimal combination of drugs for the individual patient.

With more therapeutic options becoming available over time, resistance testing is expected to play an important role in the management and treatment of disease and the development of individualized treatment regimes [see e.g. Haulbrich *et al.* JAIDS, 2001,26Sl, S51-S59].

Furthermore, the number of drug resistant diseases is also increasing. Phenotyping methodologies measure the ability of a pathogen to grow in the presence of different drugs in the laboratory. This is usually expressed as the fold change in the IC_{50} or IC_{90} values (the IC_{50} or IC_{90} value being the drug concentration at which 50% or 90% respectively of the population of pathogen is inhibited from replicating). For example, a highly resistant virus might show a 50 or even 100-fold increase in IC_{50}, for example. Some viral mutations only increase the IC_{50} by as little as 2-3 fold. On the other hand, a pathogen may exhibit hypersensitivity towards a given drug. For example, it has been demonstrated that a

combination of HIV mutations may lead to hypersensitivity of the pathogen towards a given drug.

Unlike genotyping, phenotyping is a direct measure of susceptibility, reflecting the effects and interactions of all the mutations, known or unknown, on the behaviour of the pathogen population in the presence of a drug.

The utility of drug susceptibility phenotyping is dependent on the "cut-off" value of the fold increase in, for example, the IC_{50} at which a pathogen is considered resistant. The term "cut-off value", as used herein, refers to the threshold change in susceptibility above which a pathogen is classed as having reduced susceptibility for a particular drug. Drug "resistance", as used herein, pertains to the capacity of resistance, sensitivity, susceptibility or effectiveness of the drug against the pathogen.

There has been recent debate regarding the relevance of some cut-off values currently in use. For example, for viral infections, certain groups currently use technical cut-off values, which are usually the same value for each drug-tested and are determined not by clinical criteria but, for example, by the assay variability seen on repetitive testing of a single wild type standard virus. By repeatedly running a test with the standard reference virus, the reproducibility of the test is measured and a cut-off is set at this level, (e.g., a 2.5-fold increase in IC_{50}). This provides a cut-off that depends largely on the analytical performance of the assay. This approach suffers from the limitation that it does not consider the population-based variation in drug responsiveness. In addition, such an approach does not account for different responsiveness towards different drug regimens. The limitations of setting a single cut-off for all available drugs in this way is that it tells the clinician very little about the significance of any change in susceptibility reported by a test. Indeed, some virological cut-off values are clearly out of line with known response data. For example, indications of low level resistance to non-nucleotide reverse transcriptase inhibitors (NNRTIs) does not lead to blunted responses to drugs in previously untreated individuals (Harrigan *et al.*, Bacheler *et al.*, 4th International Workshop on HIV Drug Resistance and Treatment Strategies, Sitges, Spain. Abstr. (2000)). Other assays have cut-off values that are primarily based on the reproducibility of the assay, are the same for each drug, or are not related to whether a drug might work against the pathogen in clinical practice and are, therefore, rather arbitrary.

Methods have already been described to develop more meaningful, biologically relevant cut-off values for drugs used in HIV therapy. For example, Virco measured the IC_{50} values for isolates from 1,000 untreated patients as well as many thousands of samples of HIV-1 with no resistance mutations. The average and the range of susceptibility were calculated for each drug. The cut-offs were then set at two standard deviations above the mean. This statistical term means that a test result falling above the cut-off can be said to be above the normal susceptible range with 97.5% confidence (Harrigan *et al.* World-wide variation in HIV-1 phenotypic susceptibility in untreated individuals: biologically relevant values for resistance testing. 2001. AIDS 15:1671-1677). Since the susceptibility of untreated and unmutated virus varied considerably from drug to drug, the predicted biological cut-offs are different for each drug.

The use of biological cut-offs has changed the amount of resistance being reported for HIV. For example, the biological cut-off values for the dideoxynucleoside analogues are lower than the cut-offs used previously and, in a study of 5,000 random clinical samples, revealed a higher and more realistic incidence of resistance. Conversely, the cut-offs for the non-nucleoside reverse transcriptase inhibitors are higher than those previously used.

However, although the biological cut-off values are a vast improvement to the arbitrary cut-offs used previously, there are still disparities between these predicted thresholds and the observed fold-resistance above which a clinical response is actually reduced. There is thus a great need for a method that can establish cut-off fold change resistance values that are clinically-relevant.

The present invention provides a solution to these problems, in the form of new methods for assessing the impact of pre-existing variations in drug susceptibility, whether naturally occurring or selected by previous drug exposure, on treatment response in order to establish clinically relevant cut-off values for phenotypic or genotypic resistance tests.

Summary of the invention

According to the invention, there is provided a diagnostic method for estimating for a patient the treatment response of a disease caused by a pathogen to a drug, the method comprising:

comparing the fold change resistance value of the pathogen infecting the patient to a clinical cut-off value which is the fold change resistance value at which a clinically relevant variation of clinical response is observed;

wherein the clinical cut-off value is established by modelling the clinical response of a population of patients treated with the drug to the disease caused by the pathogen as a function of the fold change resistance of the pathogen infecting the patients.

According to the invention, a threshold fold-resistance is established, above which a disease is classified as being resistant to a drug in a clinical context. The method models treatment response of the pathogen causing the disease to a particular drug as a function of baseline pathogen load, baseline resistance, baseline activity of co-administered drugs targeted to the pathogen and treatment history. By "baseline pathogen load" is meant the pathogen load of the patient measured at the start of treatment by the drug. By "baseline fold change resistance" is meant the fold change resistance to the candidate drug exhibited by the pathogen infecting the patient at the start of treatment. By "baseline activity of co-administered drugs targeted to the pathogen" is meant the activity against the pathogen of each drug administered in combination with the drug for which the treatment response is being modeled. By "treatment history" is meant the previous drug exposure of the patient (and therefore, the pathogen).

In a preferred embodiment, the cut-off value is determined as a function of treatment response data in treated subjects, considering baseline pathogen load, baseline fold change resistance, baseline activity of co-administered drugs targeted to the pathogen, and treatment history.

This method thus provides a prediction of clinical outcome at different levels of baseline resistance. According to this methodology, treatment outcome (drop in pathogen load and response rate) is modeled by drug as a function of baseline fold change resistance as determined by reference to a system that measures drug resistance phenotype or predicts drug resistance phenotype from pathogen genotype (such as VirtualPhenotype®, Virco). The models take into account effects of co-administered drugs, baseline pathogen load and, optionally, treatment history in order to avoid any bias introduced by imbalances of clinically-important characteristics. From the model, a prediction of outcome can be made at different levels of the baseline fold change resistance of the pathogen.

Using this methodology, fold change resistance values obtained by comparison of genotype with phenotype (for example, VirtualPhenotype®) are linked with clinical outcome. This is a unique approach; other research groups use different approaches whereby particular mutations or actual phenotype results are linked with clinical outcome.

The methodology of the invention is advantageous over those currently used. For example, conventional approaches do not fully account for the population-based variability in drug sensitivity. In the present method, the population may include treatment naïve and treatment experienced patients, and may be a mixed population which is not restricted to, for example, a single gender, age, race or sexual behaviour.

The method of the invention also accounts for the different responsiveness in a population towards different drugs. The drug-specific clinical cut-off values determined by this approach are more reliable parameters in estimating resistant over sensitive strains of pathogen.

The method also allows clinical cut-offs established using the method to be re-calculated depending on the type of population studied, i.e. a paediatric population may have a different clinical cut-off for a particular drug than the adult population for the same drug.

Of particular importance, this methodology allows the determination of clinical cut-offs for all marketed drugs in a uniform, scientific manner on a substantial database using data derived from response to combination therapy. Currently available cut-offs are determined by reference to a limited amount of data and may be inconsistent as they are determined using different approaches.

According to the invention, clinical cut-off values are established by modeling the clinical response of a population of patients treated with the drug to the disease caused by a particular pathogen as a function of the fold change resistance of the pathogen infecting the patients. The fold change resistance for a pathogen may be established using methods known in the art. Briefly, the sensitivity of a patient sample for a particular drug is compared with the sensitivity of a reference sample for that same drug. This may be done by a) determining the sensitivity of a patient sample for the drug; b) determining the sensitivity of a reference sample for the drug; and c) determining the patient fold change resistance from the quotient of the sensitivity obtained in step a) over the sensitivity obtained in b). Examples of preferred methods for performing these steps are described in detail in co-pending applications WO01/79540 and WO02/33402. Equivalent methods will be apparent to the person of skill in the art.

In a preferred embodiment of the invention, the cut-off fold change resistance value is calculated by reference to the log of the pathogen load drop. In such a method, a linear regression analysis is preferably performed using a set of treatment response data from

subjects harbouring the pathogen, wherein the log pathogen load drop *LogPL drop*$_i$, for the pathogen infecting a patient *i*, is modelled as the sum of all of the individual contributions for factors that influence pathogen load drop, according to the following equation:

$$LogPLdrop_i = \beta_0 + \beta_1 Log(BaselinePL_i) + \beta_2(PSS_i) + \beta_3(1/FC_i) + \varepsilon_i$$

In this equation, $BaselinePL_i$ represents the pathogen load of the patient measured at the start of treatment by the drug.

PSS_i is a phenotypic sensitivity score representing the number of active drugs in the background treatment regimen for the patient, excluding the drug whose contribution to treatment response is being modelled.

FC_i is a baseline fold change resistance.

β_0 is the intercept.

β_1 is a coefficient representing the increase in log pathogen load drop per unit increase of the log of the $BaselinePL_i$. In the case of HIV and HCV infection, *baseline PL*$_i$ is readily quantified by validated commercial assays.

β_2 is a coefficient indicating the increase in log pathogen load drop per unit increase of the number of sensitive drugs in the background treatment regimen.

β_3 is a coefficient indicating the increase in log pathogen load drop per unit increase of the inverse of FC_i. The value of this coefficient is part of the output of the described model.

ε_i, is an error term which represents the difference between the modelled prediction and the experimentally determined measurement.

PSSi, the phenotypic sensitivity score, represents the number of active drugs in the background treatment regimen for the patient, as predicted from pathogenic genotype by *Virtual*Phenotype™ or other algorithms or as measured by actual phenotype testing. The purpose of this term is to allow a drug-specific value to be extracted from treatment response data that has been collected for a patient that has received a combination of drugs. In this way, resistance data relevant solely to the particular drug under investigation is extracted. The other drugs are considered the background regimen; this may be different for different patients. It is necessary to analyze patients with different background regimens together as there would not be enough data to do a sound analysis otherwise.

During this analysis it has to be taken into account that different background regimens influence the clinical outcome in a different way. In order to do this, the activities of background drugs are summarised, by determining the number of active drugs, and thus devising a PSS (preferably judged as active according to VirtualPhenotype®). The PSS is then included in the model.

In a preferred embodiment, the PSS may be calculated based on preliminary clinical cut-offs which are determined as described. The concept of PSS is discussed in detail by DeGruttola *et al.* (Antiviral Therapy 2000; 5:41-48). In addition, the concept of continuous PSS as a variation of PSS is discussed by Bosch *et al.* (AIDS 2003, 17:1-9); Katzenstein *et al.* (AIDS 2003; 17:821-830); and Haubrich *et al.* ("Delavirdine Hypersusceptibility (DLV HS): Virological Response and Phenotypic Cut-Points – Results from ACTG 359"; 11th Conference on Retroviruses and Opportunistic Infections held on 8-11 February 2004 in San Francisco, CA, USA). The PSS may be determined by an iterative process such that the cut-off value is refined to a constant value. In subsequent iterations of the model, PSS scores based on preliminary clinical cut-offs defined in the first iteration of the model may be utilized.

FC_i, the baseline fold change, is equivalent to baseline fold change resistance. These terms are used interchangeably herein. This is a patient-specific term and is determined based on a drug susceptibility phenotype test or predicted based on the genotype of the pathogen infecting a particular patient. The phenotype exhibited by the pathogen of this genotype may be predicted in a number of ways; generally, such techniques compare the genotype to phenotype data collected from a group of patients infected with a pathogen of similar genotype. However, this does not change the fact that this fold change resistance is a characteristic of the specific pathogenic strain infecting an individual patient at baseline.

For example, prediction of baseline fold change resistance may exploit rules-based or other less direct systems of determining the drug resistance phenotype of a pathogen. An example of a less direct system is the Virtual Phenotype (Virco, Inc.; PCT/EP01/04445). Prediction of baseline fold change resistance may alternatively use other systems for determining phenotype from genotype information, such as neural networks that determine the drug resistance phenotype of a pathogen based on its genotypic information (see, for example, U. S. Patent Application No. 09/589,167; PCT/EP01/06360. The neural network

may be used to identify mutation(s) or mutation patterns that confer resistance to a drug and defines the genetic basis of drug resistance.

β_0, the intercept, is the estimated log pathogen load drop for a reference group i.e. a theoretical group of patients with a baseline pathogen load of one, an infinite fold change resistance and no sensitive drugs in the background. The purpose of this term is to improve the model fit. If it was not included, the fitted curve would be forced to pass through the origin (zero Log PL drop at zero fold change resistance), which could lead to an unrealistic model.

The error term, ε_i, represents the difference between the modelled prediction and the experimentally determined measurement i.e. the difference between the actual response of the patient and the predicted response. As more data are added to the model, additional factors that are relevant to the determination of clinical cut-off values will be added. This will improve the model fit and therefore the error of the prediction will decrease. All the β terms are estimated simultaneously by minimizing the error term.

In this methodology of this embodiment of the invention, censoring (pathogen loads beyond the assay range caused by the detection limits of pathogen load kits) affects the results and therefore procedures that take censoring into account are preferably applied. Preferably, censored values are dealt with by attempting to construct a model that is consistent from extrapolations. This model is applicable to any described methodology. Censored values are thus modelled by replacing the censored value by a maximum likelihood estimation, assuming knowledge of the standard deviation of the measurement error. For example, censored values may be dealt with using the PROC LIFEREG pre-programmed procedure in the statistical analysis package SAS that performs analyses with censored values.

An advantage of the linear regression method described above is that quantitative data about changes in pathogen load can be studied because pathogen load is considered as a continuous variable. This therefore takes into account the maximum amount of information present in the data. Estimates are corrected for covariates in the model (for example, background regimen) and therefore, do not suffer from imbalances in the covariates. Conclusions are limited to patients with covariates that are represented in the dataset in the clinical response database.

Other baseline characteristics may be added to the linear regression if relevant, resulting in the addition of new terms in the equation given above. Examples of additional baseline characteristics include the total duration of the previous treatment, and the time at which treatments were administered. For example, estimates can be corrected for duration by adding a term $\beta_4(Duration)$ in the model equation given above.

As stated above, the clinical cut-offs determine the fold change resistance with a diminished predicted clinical response to drug. In an alternative to merely classifying pathogens as sensitive or resistant, the method of this aspect of the invention preferably incorporates three classifications, namely "sensitive", associated with maximum response to drug therapy, "intermediate", associated with reduced, but still significant response to drug therapy, and "resistant", associated with little if any response to drug therapy. For example, by one set of definitions relevant for HIV response, "sensitive" may be classified as a predicted pathogen load drop of more than about 0.6 logs, "intermediate resistance" may be classified as a predicted pathogen load drop of between about 0.2 and about 0.6 logs and "resistant" may be classified as a predicted pathogen load drop of less than about 0.2 logs. In another set of definitions, "sensitive" may be classified as a predicted pathogen load drop of between about 0.5 logs and 1.0 logs. Cut-offs calculated using these definitions are highly dependant on covariates.

In a further preferred embodiment of the invention, the cut-off fold change resistance value is calculated by reference to the probability of the pathogen being susceptible to treatment by the drug for the patient, herein termed *Prob of success*. In such a method, *Prob of success* is preferably calculated by performing a logistic regression analysis using data from a clinical response dataset, wherein *Prob of success* is modelled according to the following equation:

$$Prob\ of\ success = \frac{\exp(\beta_0 + \beta_1 Log(BaselinePL_i) + \beta_2(PSS_i) + \beta_3(1/FC_i))}{(1+\exp(\beta_0 + \beta_1 Log(BaselinePL_i) + \beta_2(PSS_i) + \beta_3(1/FC_i)))}$$

The terms in the equation are the same as those described above for the embodiment of the invention described above.

This method of logistic regression does not suffer from the censoring problem described above for the linear regression model. Furthermore, the probability of success is an intuitive way of interpreting clinical outcome. One disadvantage is that by classifying the

pathogen load into successes and failures, part of the information of the continuous variable pathogen load is lost.

Estimates may also be corrected for covariates as for linear regression.

Again, like the method of the first described embodiment of the invention, the method of second described embodiment also preferably incorporates the three classifications, sensitive, intermediate resistant and resistant. On the basis that the maximum effect is defined as the treatment effect at a fold change resistance of approximately 1 fold change, and the minimum effect is defined as the treatment effect at a very high fold change resistance (i.e. when the curve reaches a plateau), the "effect range" is the difference between the maximum effect and the minimum effect. The maximum effect may be defined as the treatment effect at fold change resistance of between about 0.7 and about 1.2 fold change resistance.

Preferably, a "sensitive" genotype is classified as a predicted treatment effect of more than about 78-85% of the effect range. Preferably, "intermediate resistant" is classified as a predicted treatment effect of between about 15-25% and about 75-85% of the effect range. Preferably, "resistant" is classified as a predicted treatment effect of less than about 15-25% of the effect range. Cut-offs calculated using this method are less dependent on covariates than the method described earlier which uses predicted pathogen load drops. However, the effect range will vary for different covariates.

In a further preferred embodiment of the invention, the cut-off fold change resistance value is calculated by constructing a classification tree in order to classify the likelihood of a patient having an undetectable pathogen load after treatment with a particular drug, as a success or a failure. This methodology constructs tree-structured rules in order to classify patients as successes (undetectable pathogen load after treatment) and failures. For example, for a virus an undetectable pathogen load could be defined as a viral load of less than 400 viral copies per ml. Such a classification tree has the advantage that it is very visual and easy to interpret, although it suffers from the limitation that the decisions do not take into account the value of certain other relevant parameters. Imbalances for such parameters may therefore influence the decision taken for a certain parameter. However, such trees provide insights into the importance of several parameters and this can be helpful in the fitting process of the linear regression and logistic regression approaches described above.

The classification tree poses queries, in which the answer to each query results in either the left or the right branch of the tree being taken at each stage. For example, the first query may preferably consider the fold change resistance of the pathogen genotype to the drug in question e.g. is fold change for the drug TDF (tenofovir) < 1.35? If yes, the left branch is taken, if no the right branch is taken. As with the methods of the aspects of the invention described previously, the other factors queried include the log baseline pathogen load and the phenotypic sensitivity score. The numbers at the termini of the final branches represent the response rate (1 = 100% response). An example of a classification tree according to the invention is provided in Figure 8.

In this embodiment of the invention, the clinical cut-off is defined as the fold change resistance threshold value that makes the best distinction between successful and unsuccessful treatments i.e. the most suitable value posed in the query that bifurcates the tree into the left and right branches. The population is thus split into two subgroups: one with a high success rate and one with a low success rate. The clinical cut-off is chosen as the fold change that makes the difference between the two groups as large as possible.

Preferably, two or all three of the methods of the above-described embodiments of the invention are performed for each dataset and candidate drug. The clinical cut-offs can in this manner be calculated for each of the approaches. From the analysis results, the most appropriate values for lower and higher cut-offs are selected, taking into account the advantages and the disadvantages of the separate approaches. This selection will only be made if the results of the approaches are consistent or if possible inconsistencies can be explained. If there are unexplained inconsistencies between the results, it can be concluded that more data need to be gathered before a clinical cut-off can be determined.

For example, if the results of the different approaches are consistent (preferably clinical cut-off difference < 0.5) then the predictions are deemed to be consistent. If the results differ more than that, the disparities need to be explained. For example, if we suppose that the population contains 90% censored values and the linear regression gives a clinical cut-off of 0.9 while the logistic regression gives a clinical cut-off of 3.5, then in this example the linear regression results are less reliable because too much correction has to be made for censoring and there is too little information contributed by "complete" observations.

The methods of the invention can be repeated for each possible drug or therapeutic agent known or suspected to be associated with disease resistance, or towards which a resistance

can be expected to appear. As such, according to another embodiment of the invention, the clinical cut-offs generated can be presented as a list of cut-offs against or in respect of individual drugs or individual therapeutic agents, for each pathogen.

As used herein, the term "drug" includes, but is not limited to, a pharmaceutical, bactericide, fungicide, antibiotic, or anticancer, antiviral, anti-bacterial anti-fungal, anti-parasitical or any other compound or composition that can be used in therapy or therapeutic treatment.

A "patient" may be any organism, particularly a human or other mammal, suffering from a disease or in need or desire of treatment for a disease. A patient includes any mammal, including farm animals or pets, and includes humans of any age or state of development. A group of patients useful to establish treatment response as a function of the distribution of fold change resistances may be as low as 10 to 50 patients, 50 to 500 patients, or, more preferably, will comprise a population of 500 or more patients. The distribution fold change resistances can be a normal distribution (Gaussian distribution) or can be a non-normal distribution. The non-normal distribution may be transformed to obtain a normal distribution.

The patient samples may be from treatment naïve or treatment experienced subjects, with or without resistance to one or more drugs.

As used herein, the term "disease" refers to a disease caused by infection with a pathogen. The term "pathogen", as used herein, is used broadly and refers not just to pathogenic microorganisms, but includes any disease-causing agent. Examples include bacteria, viruses such as human immunodeficiency virus (HIV), hepatitis C (HCV) or hepatitis B (HBV), prions, algae, fungi, protozoa and malignant cells. This invention is particularly useful for viral diseases such as HIV.

A "patient sample" is herein defined as any sample obtained from an individual suffering from or predicted to be suffering from a disease caused by a pathogen, and includes tissues such as blood, serum plasma, urine, saliva, semen, breast milk, faeces, mucous samples, cells in cell culture, cells which may be further cultured, biopsy samples and so on. In one embodiment, for a patient infected with HIV, any biological sample-containing virus may be used. Of this patient sample, the pathogen itself may be used or alternatively a protein, or nucleic acid derived from the pathogen. Preferably, the pathogen is a virus, such as a retrovirus. Preferably the biological sample contains a virus chosen from HIV, HCV

(Hepatitis C Virus) and HBV (Hepatitis B virus). In another embodiment, for a cancer patient, the patient sample may contain cells, tissue cells, mutated cells, malignant cells, cancer cells, whole or partial tumours, biopsy tissue, etc. Preferably, the pathogen is a malignant cell. A "reference sample" is defined as a standard laboratory reference pathogen such as, for example, in the case of HIV, the HIV LAI IIIB strain. One strain generally used as the reference "wild type" sequence for HIV is HXB2. This viral genome comprises 9718 bp and has an accession number in Genbank at NCBI M38432 or K03455 (gi number: 327742). Reference or wild type sequences for use in the invention in the field of specific diseases, infections or diseases caused by specific pathogens can be easily obtained from publicly available databases.

"Susceptibility" or "sensitivity" to a drug refers to the capacity of the disease, and/or pathogen to be affected by the drug. "Resistance" refers to the degree to which the disease and/or pathogen is unaffected by the drug. The sensitivity, susceptibility or resistance of a disease towards a drug may be expressed by means of an IC_{50} value. The IC_{50} value is the concentration at which a given drug results in a reduction of the pathogen's growth compared to the growth of the pathogen in the absence of a drug. Resistance of a disease to a drug may be caused by alterations in phenotype or genotype. Genotypic alterations include mutations, single nucleotide polymorphisms, microsatellite variations, and/or epigenetic variations such as methylation. Phenotypic variations may be effected by genotypic variations or by post-translational modification.

Any method capable of measuring changes in the ability of a pathogen to grow in the presence of a drug(s) can be used in the method of the present invention. Such methods of phenotyping include all methods known to persons of skill in the art. Known genotyping methods may also be applicable.

For example, and by way of illustration, methods for phenotyping bacteria suitable for use in the present invention include, but are not limited to, measurement of inhibitory zone diameters (see, e.g., Guoming *et al.*, Sex Transm. Dis. 27 (2) : 115-8 (2000)), colorimetric indicator methods (see, e.g., Lozano-Chiu *et al.*, Diagn Microbiol Infect Dis. 1998 Jul;31(3):417-24), and broth macrodilution method (see, e.g., Iwen *et al.*, J. Clin. Microbiol. 34 (7) : 1779-83 (1996)).

As an additional illustrative example, methods for phenotyping pathogens suitable for use in the present invention include, but are not limited to, plaque reduction assays, PBMC p24

growth inhibition assays (see, e.g., Japour *et al.*, Antimicrob Agents Chemother. 1993 May;37(5):1095-101; Kusumi *et al.*, J. Virol. 66 : 875-885 (1992)), recombinant virus assays (see, e.g., Kellam & Larder, Antimicrob. Agents Chemother. 38 : 23-30 (1994); and Pauwels *et al.*, 2nd International Workshop on HIV Drug Resistance and Treatment Strategies, Lake Maggiore, Italy. Abstr. 51 (1998)); the use of GFP as a marker to assess the susceptibility of anti-viral inhibitors (Marschall *et al.*, Institute of Clin. and Mol. Virol., University of ErlangerNuremberg, Schlobgarten, Germany); and cell culture assays (Hayden *et al.*, N. Eng. J. Med. 321 : 1696-702 (1989)).

Though the invention may be used with any phenotype or genotype measuring test or assay that determines resistance, the following descriptions are designed to describe further possible applications of the invention.

In one embodiment, the clinical cut-off values may be used in concert with direct phenotype assays, for example, Antivirogram™ (Virco, Inc.; WO 97/27480, US 6,221,578). This assay is a phenotypic resistance assay that measures, in controlled laboratory conditions, the level of resistance of the HIV derived from an individual patient to each of the anti-HIV drugs currently available. The resistant "behaviour" of the virus may be the combined result of the effects of many different mutations and the complex interactions between them, including genetic changes that have not even been identified yet. In other words, it is a direct measure of resistance.

The test provides a quantitative measure of viral resistance to all the available drugs. This is expressed in terms of the IC_{50}. This is then compared to the IC_{50} for fully sensitive, non-mutated "wild-type" virus. The resistance of the sampled virus to each drug is then expressed in terms of a fold-change in IC_{50} compared to wild type.

The addition of "clinical cut-offs", as described in this application, to the report enables physicians to identify the drug(s) that are no longer clinically active and helps in the selection of the optimal combination of drugs for the individual patient. In one embodiment, the method of the present invention concerns a diagnostic tool for determining the resistance of a patient to at least one HIV drug comprising the clinical cut-off fold change resistance value for said at least one drug as determined herein. The diagnostic tool includes phenotypic resistance tests such as the Antivirogram®, VirtualPhenotyping® and Phenosense.

The invention includes methods to determine resistance towards HIV compounds such as suramine, pentamidine, thymopentin, castanospermine, dextran (dextran sulfate), foscarnet-sodium; nucleoside reverse transcriptase inhibitors, e.g. zidovudine, didanosine, zalcitabine or lamivudine, stavudine, abacavir and the like; non-nucleoside reverse transcriptase inhibitors such as nevirapine, efavirenz, delavirdine, 4-[4-(2,4,6-trimethyl-phenylamino)-pyrimidin-2-ylamino]-benzonitrile, and the like; phosphonate reverse transcriptase inhibitors, e.g. tenofovir and the like; compounds of the TIBO (tetrahydro-imidazo[4,5,1-jk][1,4]-benzodiazepine-2(1*H*)-one and thione)-type e.g. (S)-8-chloro-4,5,6,7-tetrahydro-5-methyl-6-(3-methyl-2-butenyl)imidazo-[4,5,1-jk][1,4]benzodiazepine-2(1*H*)-thione; compounds of the α-APA (α-anilino phenyl acetamide) type e.g. α-[(2-nitrophenyl)amino]-2,6-dichlorobenzene-acetamide and the like; inhibitors of trans-activating proteins, such as TAT-inhibitors, e.g. RO-53335, or REV inhibitors, and the like; protease inhibitors *e.g.* indinavir, ritonavir, saquinavir, lopinavir (ABT-378), nelfinavir, amprenavir, (3-[(4-amino-benzenesulfonyl)-isobutyl-amino]-1-benzyl-2-hydroxy-propyl)-carbamic acid hexahydro-furo[2,3-b]furan-3-yl ester, 1-benzyl-2-hydroxy-3-[isobutyl-(4-methoxy-benzenesulfonyl)-amino]-propyl)-carbamic acid hexa-hydrofuro[2,3-b]furan-3-yl ester, BMS-232632, VX-175 and the like; fusion inhibitors, e.g. T-20, T-1249 and the like; CXCR4 receptor antagonists, e.g. AMD-3100 and the like; inhibitors of the viral integrase; ribonucleotide reductase inhibitors, e.g. hydroxyurea and the like; and those compounds disclosed in WO99/67417, EP-A-945443 and WO00/27825. Other examples of drugs will be well known to those of skill in the art.

In one embodiment the effect of drugs on HBV may be monitored using technologies such as disclosed by Isom *et al.* (WO 99/37821, Delaney *et al.* Antimicrob. Agents Chemotherap. 2001, 45 (6) 1705-1713).

In one embodiment the effect of drugs on HCV towards therapy may be determined using techniques such as described by Rice (WO 97108310, WO 98/39031) and Barthenschlager (EP 1043399).

The primary aim of the invention is to predict the resistance of a disease to a particular drug. In addition, however, the invention encompasses methods of evaluating currently applied drugs and thus monitoring these drugs with a view to assessing the effectiveness of that drug and proposing alternative drug(s) or optimizing the drug if deemed appropriate. Such methods involve obtaining a sample containing a disease-causing pathogen from a

patient, and then performing the steps described in any one or more of the embodiments of the invention described above.

It will be apparent to the skilled reader that while the invention has been described in the below examples with respect to viruses, particularly HIV, the present invention has broad applicability to any disease state where it is desired to correlate genotypic information with phenotypic profiles and assess the threshold at which a fold change resistance is clinically significant. One skilled in the art could readily take the following discussion of the invention with the HIV virus and through the exercise of routine skill apply this invention to other diseases (such as other viral infections, malignant cells, cancer, bacterial infections, other pathogens, and the like) to correlate genotypic information to predict phenotypic response, assess drug resistance, and eventually develop a treatment regime of drugs for a particular patient. One skilled in the art will also know that many virus species comprise many strains; for instance, HIV comprises HIV-2 in addition to HIV-1 and both groups are further divided into groups (such as groups O and M for HIV-1).

The above methods are diagnostic methods. Further aspects of the invention provide diagnostic kits for performing any one of the diagnostic methods of the invention described above. The invention further relates to a diagnostic system as herein described for use in any of the above described methods.

According to yet another embodiment, the present invention relates to a diagnostic system for predicting clinical response to a drug of a disease causing pathogen comprising: a) means for obtaining a genetic sequence of the disease producing pathogen; b) means for identifying at least one mutation in the genetic sequence of the disease producing pathogen; c) genotype database means comprising genotype entries; d) phenotype database means comprising phenotypes of patient fold change response values; e) clinical response database means comprising clinical response to drug treatment data for reference sample patients; f) correlation means correlating a genotype entry with a phenotype, where the genotype entry corresponds with the obtained genetic sequence of the disease producing pathogen; g) means for modeling clinical response to a drug of the disease causing pathogen by determining whether the patient fold change response is above a cut-off value, wherein the cut-off value is determined using the clinical response database means and comprises the fold change response value at which a clinically relevant diminished clinical

response is observed; and h) means for predicting the clinical response to a drug of a disease by determining whether the patient fold change response is above the cut-off value.

As described above, the cut-off value is determined as a function of treatment response data in treated subjects, considering baseline pathogen load, baseline fold change resistance, baseline activity of co-administered drugs targeted to the pathogen, and treatment history. The means for predicting the resistance are preferably computer means.

A still further aspect of the invention relates to a computer apparatus or computer-based system adapted to perform any one of the methods of the invention described above.

In a preferred embodiment of the invention, said computer apparatus may comprise a processor means incorporating a memory means adapted for storing data; means for inputting data relating to the genotype exhibited by a particular disease causing pathogen; and computer software means stored in said computer memory that is adapted to perform a method according to any one of the embodiments of the invention described above and output a prediction of the resistance of a disease causing pathogen toward a drug.

A computer system of this aspect of the invention may comprise a central processing unit; an input device for inputting requests; an output device; a memory; and at least one bus connecting the central processing unit, the memory, the input device and the output device. The memory should store a module that is configured so that upon receiving a request to model the response to a drug of a disease causing pathogen, it performs the steps listed in any one of the methods of the invention described above.

In the apparatus and systems of these embodiments of the invention, data may be input by downloading the data from a local site such as a memory or disk drive, or alternatively from a remote site accessed over a network such as the internet. Data may be input by keyboard, if required.

The generated results may be output in any convenient format, for example, to a printer, a word processing program, a graphics viewing program or to a screen display device. Other convenient formats will be apparent to the skilled reader.

The means adapted to predict the resistance of a disease causing agent to a drug will preferably comprise computer software means. As the skilled reader will appreciate, once the novel and inventive teaching of the invention is appreciated, any number of different computer software means may be designed to implement this teaching.

According to a still further aspect of the invention, there is provided a computer program product for use in conjunction with a computer, said computer program comprising a computer readable storage medium and a computer program mechanism embedded therein, the computer program mechanism comprising a module that is configured so that upon receiving a request to predict the resistance of a disease to a drug, it performs the steps listed in any one of the methods of the invention described above.

The invention further relates to systems, computer program products, business methods, server side and client side systems and methods for generating, providing, and transmitting the results of the above methods.

The invention will now be described by way of example with particular reference to a specific system that implements the process of the invention. As the skilled reader will appreciate, variations from this specific illustrated embodiment are of course possible without departing from the scope of the invention.

Brief description of the Figures

Figure 1: Example of the structure of a clinical data base used in the present invention.

Figure 2: Example of linear regression curve showing censored and uncensored observations, where log viral load drop is modelled as a function of baseline fold change resistance.

Figure 3: Example of linear regression curve for TNF, where log viral load drop is modelled as a function of baseline fold change resistance and a first definition of clinical cut-off is applied.

Figure 4: Example of linear regression curve for TNF, where log viral load drop is modelled as a function of baseline fold change resistance and a second definition of clinical cut-off is applied.

Figure 5: Example of logistic regression curve for TNF, where probability of failure is modelled as a function of baseline fold change resistance and a second definition of clinical cut-off is applied.

Figure 6: Example of linear regression curve for TNF, where log viral load drop is modelled as a function of baseline fold change resistance and a third definition of clinical cut-off is applied.

Figure 7: Example of logistic regression curve for TNF, where probability of failure is modelled as a function of baseline fold change resistance and a third definition of clinical cut-off is applied.

Figure 8: Example of classification tree for TNF. This gives results of the same order as the linear and logistic regression methodologies.

Example: Process description of the determination of clinical cut-offs

Step 1: Clinical Data Base

Databases of studies for patients with tenofovir containing regimens and consisting of patient baseline demographic characteristics, clinical outcome results with viral load and resistance data (Fold change), were retrieved and remapped according to a common structure allowing a meta-analysis. The structure consisted of viral load data set, viral load measurements and sampling dates, CD4+ data set which contains CD4+ counts and sampling dates, resistance data set containing the fold changes to different antivirals and sampling dates; patient data set with patient information such as age, gender, race, treatment history; treatment data set with drug regimens, start and stop dates, doses, formulations, frequency of intake. The structure of such a clinical data base can be seen in figure 1.

Step 2: Modelling

The clinical outcome results (drop in viral load and response rate) were modelled in function of baseline fold change (FC) as determined by virtual phenotype (see WO01/79540 and WO02/33402; also http://www.vircolab.com). The models applied were linear regression, logistic regression, and a classification tree. These models also took into account effects of the concomitant HIV drugs (PSS), baseline viral load (Baseline $Log(Vl)_i$) and, optionally, treatment history in order to avoid bias introduced by imbalances of important characteristics. From the models, a prediction of clinical outcome could be made at different levels of the baseline fold change resistance.

In the linear regression model, the proposed equation was the following:

$$\text{LogVL drop}_i = \beta 0 + \beta 1 * \text{Baseline Log(Vl)}_i + \beta 2 * \text{PSS}_i + \beta 3 * (1/\text{FC})_i + \xi_i$$

where i represented the patient, β0 the intercept, β1, β 2 and β3 coefficients indicated the increase in log viral load drop per unit increase of respectively the baseline log VL, the number of sensitive drugs in the background regimen and the inverse of the baseline fold change. ξ_i was a random error term indicating the deviation of the patient from the value predicted by the model. Interactions between all the factors were evaluated and other baseline characteristics, i.e treatment history, were added if relevant. After applying the regression model, the curve as depicted in Figure 2 was obtained.

In the logistic regression, the proposed equation was the following:

$$\textit{Prob of success} = \frac{\exp(\beta_0 + \beta_1 Log(BaselineVL_i) + \beta_2(PSS_i) + \beta_3(1/FC_i))}{(1+\exp(\beta_0 + \beta_1 Log(BaselineVL_i) + \beta_2(PSS_i) + \beta_3(1/FC_i)))}.$$

where β1, β2, and β3 represented the log odds ratio of success for the corresponding factors in the model. After applying the logistic regression model, the curves as depicted in Figures 5 and 7 were obtained. In the classification trees model, tree-structured rules were constructed in order to classify patients in successes (undetectable viral load after treatment) and failures. The same parameters as for the other techniques were considered. The tree shown in Figure 8 was obtained after applying the classification tree model.

When viral load results under the detection limits are obtained, biases could be introduced if the detection limit values are considered when calculating viral load drops and using those in the linear regression model. To avoid this, censoring needed to be taken into account and therefore the PROC LIFEREG facility in the SAS package was employed.

An advantage of this regression model is that it takes into account the maximum amount of information present in the data, i.e. correlating specific clinical responses with specific Fold changes while the other two models clusters the patients in two groups (successes versus failures), thus not taking into account differences in responses within the same group. Estimates are corrected for covariates in the model (e.g. background regimen) and therefore, they do not suffer from imbalances in the covariates. Conclusions are limited to patients with covariates that are represented in the clinical database.

Logistic regression does not suffer from the censoring problem and the probability of success is an intuitive way of interpreting clinical outcome. However, by binning the viral

load into successes and failures, part of the information of the continuous variable viral load is lost. Estimates are also corrected for covariates as for linear regression.

Classification trees are very visual and easy to interpret, but they have the disadvantage that the decisions do not take into account the value of other relevant parameters. This implies that imbalances for other parameters may influence the decision taken for a certain parameter. However, they provide insights in the importance of several parameters and this can be helpful in the fitting process of the other approaches. Figure 8 shows an example of a classification tree.

Step 3: Determination of the clinical cut-off.

Clinical responses were predicted in the models developed in previous step 2. In order to determine the fold changes at which clinically relevant diminished clinical responses can be observed, three definitions of clinical cut-offs were considered:

Definition 1:

Sensitive: predicted viral load drop is more than 0.6 logs.

Intermediate resistant: predicted viral load drop is between 0.2 and 0.6 logs.

Resistant: predicted viral load drop is less than 0.2 logs.

Definition 2:

The maximum effect was defined as the treatment effect at fold change 1, and the minimum effect was defined as the treatment effect at a very high fold change (i.e. when the curve reached a plateau). The effect range was then the difference between the maximum effect and the minimum effect.

Sensitive: the predicted treatment effect is more than 80% of the effect range.

Intermediate resistant: the predicted treatment effect is between 20% and 80% of the effect range.

Resistant: the predicted treatment effect is less than 20% of the effect range.

Definition 3:

Definition 3 was a variant of definition 2. The lower cut-off was defined as the fold change that most optimally distinguished patients between successful and unsuccessful treatments.

The methodology was applied for Tenofovir on a population taking two active drugs besides tenofovir and with a baseline Log(Vl) of 4.

When we applied definition 1 on the linear regression model, the observed drop in log viral load was-0.6 at fold change 3.73 (fig 3). No higher cut-off could be derived as this population experienced a drop in Log(VL) greater then 0.2 even with a high baseline fold change for tenofovir. This could be explained by the effect of the active background regimen in this population.

When we applied definition 2 on the linear regression model (fig 4), the observed drop in log viral load was -1.48 at fold change 1, and –0.28 at the assymptotic fold change. Therefore the effect range was $-0.28 + 1.48 = 1.2$.

20% of this effect range was observed at fold change 5 (and this value was considered as the upper clinical cut-off value).

80% of the effect range was observed at fold change 1.25 (and this value was considered as the lower clinical cut-off value).

To predict the resistance according to this regression model, we determined whether the patient fold change resistance was above, below, or in between the clinical cut-off as calculated according to definition 2. So, when the FC of patient was of 0.8 (below the lower clinical cut-off), a normal clinical response was predicted. If the FC of the patient was of 2 (above the lower clinical cut-off and below the upper clinical cut-off), a reduced clinical response was predicted. If the FC was of 7 (above the clinical cut-off), then the clinical response was predicted as being minimal.

Definition 2 was also applied to the logistic model (fig. 5) and this resulted in a lower cut-off at 1.2 FC and a higher cut-off at 3.81 FC.

The results for definition 3 are depicted in fig 6 and 7.

The Tenofovir results for the population with 2 active drugs in the regimen and a baseline Log(VL) of 4 are summarized in the Table below.

Definition of clinical cut-off	Population	Properties of the subgroup	*Linear Regression*		*Logistic Regression*		*Classification Tree*	
			Lower CO	Higher CO	Lower CO	Higher CO	Lower CO	Higher CO

Definition 1	*Subgroup 1*	*PSS=2, baseline Log(VL) = 4*	3.73	> assay limit	NA	NA	NA	NA
	Subgroup 2	*PSS=0, baseline Log(VL) = 4*	1.68	3.8	NA	NA	NA	NA
	Subgroup 3	*PSS=2, baseline Log(VL) = 5*	> assay limit	> assay limit	NA	NA	NA	NA
	Overall		NA	NA	NA	NA	NA	NA
Definition 2	*Subgroup 1*	*PSS=2, baseline Log(VL) = 4*	1.25	5	1.2	3.81	NA	NA
	Subgroup 2	*PSS=0, baseline Log(VL) = 4*	1.25	5	1.16	3.36	NA	NA
	Subgroup 3	*PSS=2, baseline Log(VL) = 5*	1.25	5	1.17	3.4	NA	NA
	Overall		1.25	5	NA	NA	NA	NA
Definition 3	*Subgroup 1*	*PSS=2, baseline Log(VL) = 4*	1.1	5	1.2	3.81	1.15	NA
	Subgroup 2	*PSS=0, baseline Log(VL) = 4*	1.1	5	1.2	3.36	1.15	NA
	Subgroup 3	*PSS=2, baseline Log(VL) = 5*	1.1	5	1.2	3.4	1.15	NA
	Overall		1.1	5	1.2	NA	1.15	NA

NA: Not Applicable

From the Table, it can be derived that the lower cut-off for definition is 1.2 and the higher cut-off ranges from 3.81 to 5 for the population of patients taking 2 active drugs besides tenofovir and with a baseline log(VL) of 4. The variation in cut-offs determined by the different definitions is a result of the different influence of the covariates such as PSS and log VL. That is, the influence of the covariates is significant when using definition 1 and less significant when using definition 2.

Definition 1 can only be applied on the linear regression model. The clinical cut-offs determined using definition 1 are highly dependent on the characteristics of the subpopulation. This is due to the fact that definition 1 describes the potency of the whole drug regimen while definition 2 is related only to the activity of the drug under consideration and its resistance profile. In other words, the activity of the background regimen together with the drug under investigation determines the viral load drop that the patient will experience and hence the dependence of the cut-off on the background regimen. The activity of the background regimen does not change the resistance profile in a profound way, therefore the clinical cut-offs do not vary considerably with the population characteristics.

Step 4: Validation of the cut-offs

The models were validated using bootstrapping and repeating the steps described above several times. Bootstrapping is a resampling technique in which pseudo-populations of the

same size as the original population are created by randomly drawing samples from the original population. Analysis of each of these populations gives a sense of the sampling variability of the clinical cut-off.

The problem is tackled from different points of view in order to assess the robustness of the analysis results. The clinical cut-offs obtained could be further refined by adding more data sets and by taking more characteristics of the patients into account.

CLAIMS

1. A diagnostic method for estimating for a patient the treatment response of a disease caused by a pathogen to a drug, the method comprising:

comparing the fold change resistance value of the pathogen infecting the patient to a clinical cut-off value which is the fold change resistance value at which a clinically relevant variation of clinical response is observed;

wherein the clinical cut-off value is established by modeling the clinical response of a population of patients treated with the drug to the disease caused by the pathogen as a function of the fold change resistance of the pathogen infecting the patients.

2. A method according to claim 1, wherein the cut-off value is determined as a function of treatment response data in treated subjects, considering baseline pathogen load, baseline fold change resistance and baseline activity of coadministered drugs targeted to the pathogen.

3. A method according to claim 1, wherein the cut-off value is calculated by reference to the pathogen load drop.

4. A method according to claim 3, wherein the cut-off value is calculated by reference to the log pathogen load drop.

5. A method according to claim 4, wherein the log pathogen load drop is calculated by performing a linear regression analysis using data from a dataset of treatment response data, wherein the log pathogen load drop $LogPL\ drop_i$, for the pathogen infecting a patient i, is modelled as the sum of all of the individual contributions for factors that influence pathogen load drop, according to the following equation:

$$LogPLdrop_i = \beta_0 + \beta_1 Log(BaselinePL_i) + \beta_2(PSS_i) + \beta_3(1/FC_i) + \varepsilon_i$$

wherein $BaselinePL_i$ represents the pathogen load of the patient measured at the start of treatment by the drug,

PSS_i is a phenotypic sensitivity score representing the number of active drugs in the background treatment regimen for the patient, excluding the drug whose contribution to treatment response is being modelled,

FC_i is a baseline fold change resistance,

β_0 is the intercept,

β_1 is a coefficient representing the increase in log pathogen load drop per unit increase of the log of the $BaselinePL_i$,

β_2 is a coefficient indicating the increase in log pathogen load drop per unit increase of the number of sensitive drugs in the background treatment regimen,

β_3 is a coefficient indicating the increase in log pathogen load drop per unit increase of the inverse of FC_i,

and wherein the error term, ε_i, represents the difference between the modelled prediction and the experimentally determined measurement.

6. A method according to claim 1, wherein the cut-off response value is calculated by reference to the probability of the pathogen being susceptible to treatment by the drug for the patient, herein termed *Prob of success*.

7. A method according to claim 6, wherein *Prob of success* is calculated by performing a logistic regression analysis using data from a dataset of treatment response data, wherein *Prob of success* is modelled according to the following equation:

$$Prob\ of\ success = \frac{\exp(\beta_0 + \beta_1 Log(BaselinePL_i) + \beta_2(PSS_i) + \beta_3(1/FC_i))}{(1+\exp(\beta_0 + \beta_1 Log(BaselinePL_i) + \beta_2(PSS_i) + \beta_3(1/FC_i)))}$$

wherein $BaselinePL_i$ represents the pathogen load of the patient measured at the start of treatment by the drug,

PSS_i is a phenotypic sensitivity score representing the number of active drugs in the background treatment regimen for the patient, excluding the drug whose contribution to treatment response is being modelled,

FC_i is a baseline fold change resistance,

β_0 is the intercept,

β_1 is a coefficient representing the increase in log pathogen load drop per unit increase of the log of the $BaselinePL_i$,

β_2 is a coefficient indicating the increase in log pathogen load drop per unit increase of the number of sensitive drugs in the background treatment regimen, and

β_3 is a coefficient indicating the increase in log pathogen load drop per unit increase of the inverse of FC_i.

8. A method according to claim 1, wherein the cut-off fold change resistance value is calculated by reference to the likelihood of a patient achieving treatment success or

failure, where a definition of success is having an undetectable pathogen load after treatment with a particular drug, using a classification tree.

9. A method according to claim 8, wherein the clinical cut-off is defined as the fold change resistance threshold value that makes the best distinction between the population with successful treatments and the population with unsuccessful treatments.

10. A method according to any one of the preceding claims, wherein the baseline fold change resistance is determined by comparing the genotype of the disease causing pathogen to phenotype data collected from a group of patients infected with a pathogen of similar genotype.

11. A method according to claim 10, wherein the baseline fold change resistance is determined using the Virtual Phenotype system, or a variation thereof.

12. A method according to claim 1, that incorporates two or more of the methods recited in claims 5, 7 and 8.

13. A method according to any one of the preceding claims which is a computer-implemented method.

14. A method according to claim 13, which is an automated method.

15. A method according to any one of the preceding claims, wherein the disease causing pathogen is obtained from a patient sample chosen from a blood sample, a biopsy sample, a plasma sample, a saliva sample, a tissue sample, and a bodily fluid or mucous sample.

16. A method according to any one of the preceding claims, wherein the disease causing pathogen is a virus.

17. A method according to claim 16, wherein the disease causing virus is chosen from HIV, HCV and HBV.

18. A method according to any one of the preceding claims, wherein the method is performed for a number of candidate drugs so as to provide information on the predicted fold resistance exhibited by the pathogen to a spectrum of candidate drugs.

19. A diagnostic method for optimising a drug therapy in a patient, comprising performing a method according to any one of the preceding claims for each drug or combination of drugs being considered to obtain a series of drug resistance phenotypes and therefore assess the effect of the plurality of drugs or drug combinations on the pathogen with which the patient is infected and selecting the drug or drug combination for which the pathogen is predicted to have the lowest fold resistance.

20. Use of a method according to any one of the preceding claims for assessing the efficiency of a patient's therapy or for evaluating or optimizing a therapy.

21. A diagnostic system for predicting clinical response to a drug of a disease causing pathogen comprising: a) means for obtaining a genetic sequence of the disease producing pathogen; b) means for identifying at least one mutation in the genetic sequence of the disease producing pathogen; c) genotype database means comprising genotype entries; d) phenotype database means comprising phenotypes of patient fold change response values; e) clinical response database means comprising clinical response to drug treatment for reference sample patients; f) correlation means correlating a genotype entry with a phenotype, where the genotype entry corresponds with the obtained genetic sequence of the disease producing pathogen; g) means for modelling clinical response to a drug of the disease causing pathogen by determining whether the patient fold change response is above a cut-off value, wherein the cut-off value is determined using the clinical response database means and comprises the fold change response value at which a clinically relevant diminished clinical response is observed; and h) means for predicting the clinical response to a drug of a disease by determining whether the patient fold change response is above the cut-off value.

22. A diagnostic system according to claim 21, wherein the cut-off value is determined as a function of treatment response data in treated subjects, considering baseline pathogen load, baseline fold change resistance, baseline activity of co-administered drugs targeted to the pathogen and treatment history.

23. A computer apparatus or computer-based system adapted to perform the method of any one of claims 1-18.

24. A computer program product for use in conjunction with a computer, said computer program comprising a computer readable storage medium and a computer program mechanism embedded therein, the computer program mechanism comprising a module that is configured so that upon receiving a request to predict the response of a disease caused by a pathogen to a drug it performs a method according to any one of claims 1-18.

Figure 1:

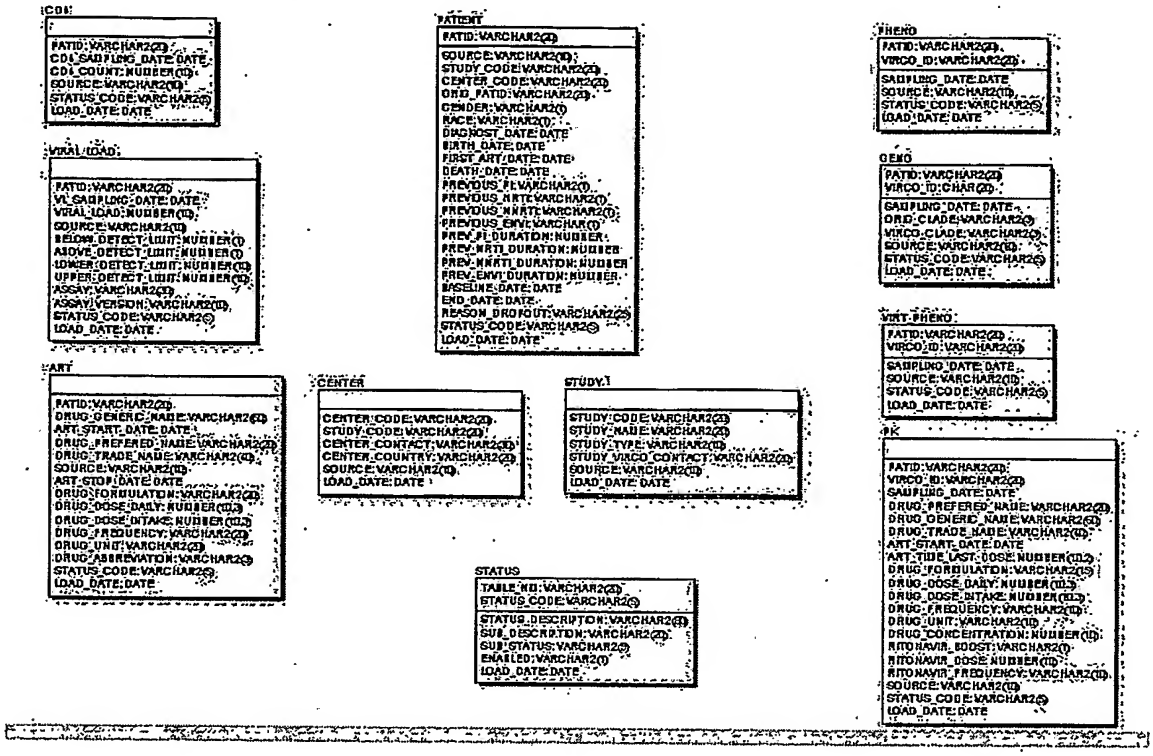

Figure 2

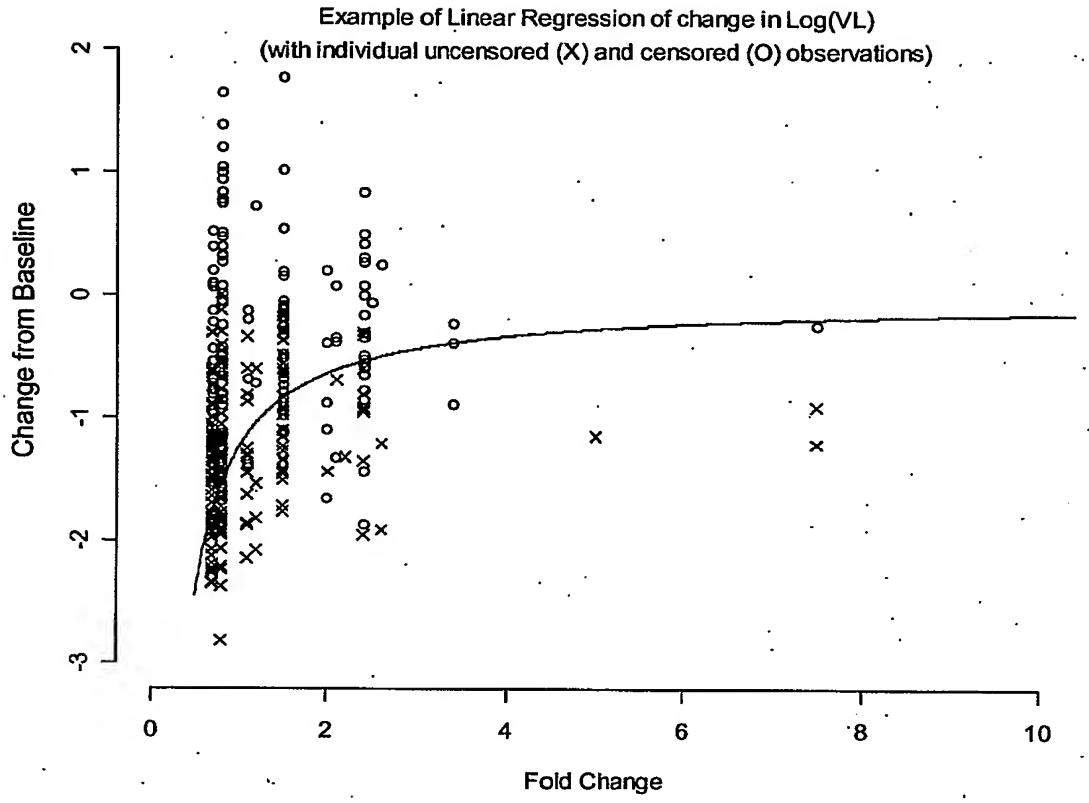

Figure 3

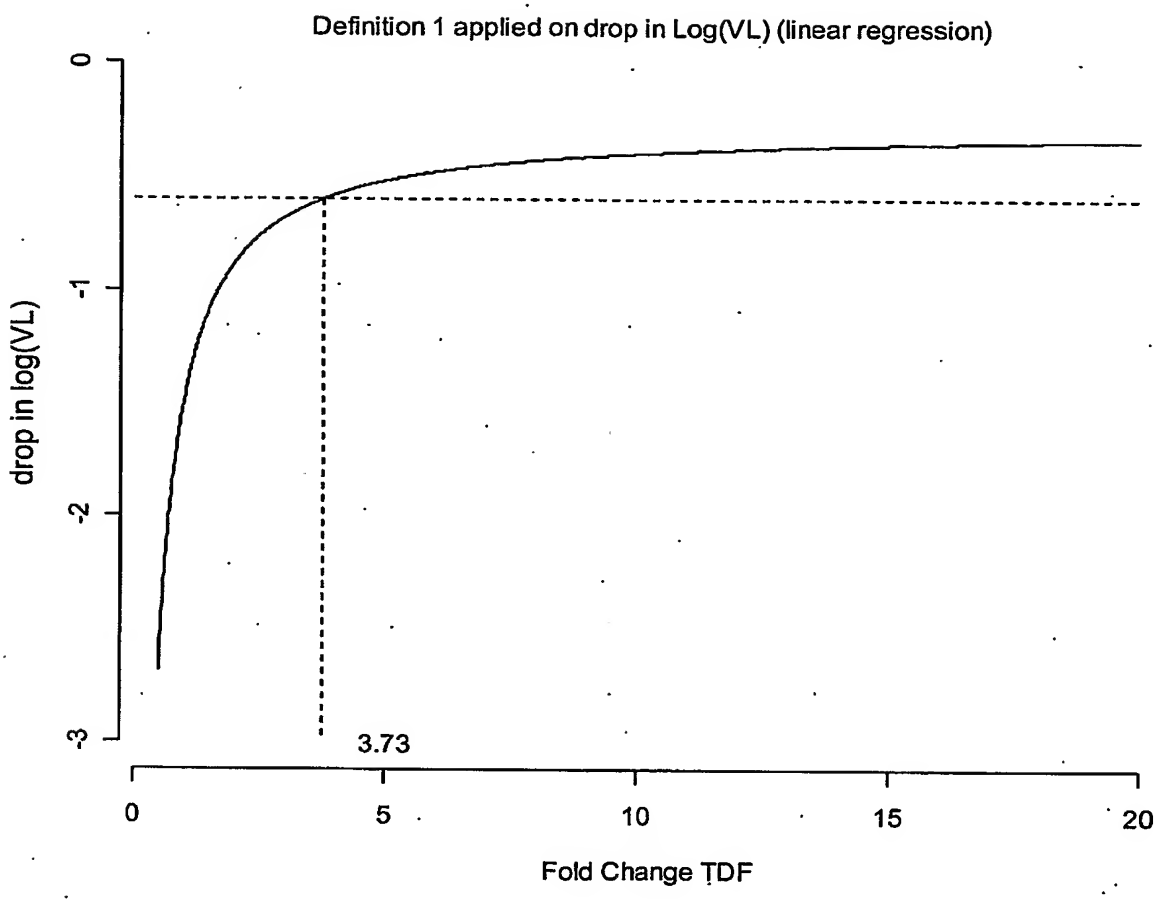

Figure 4

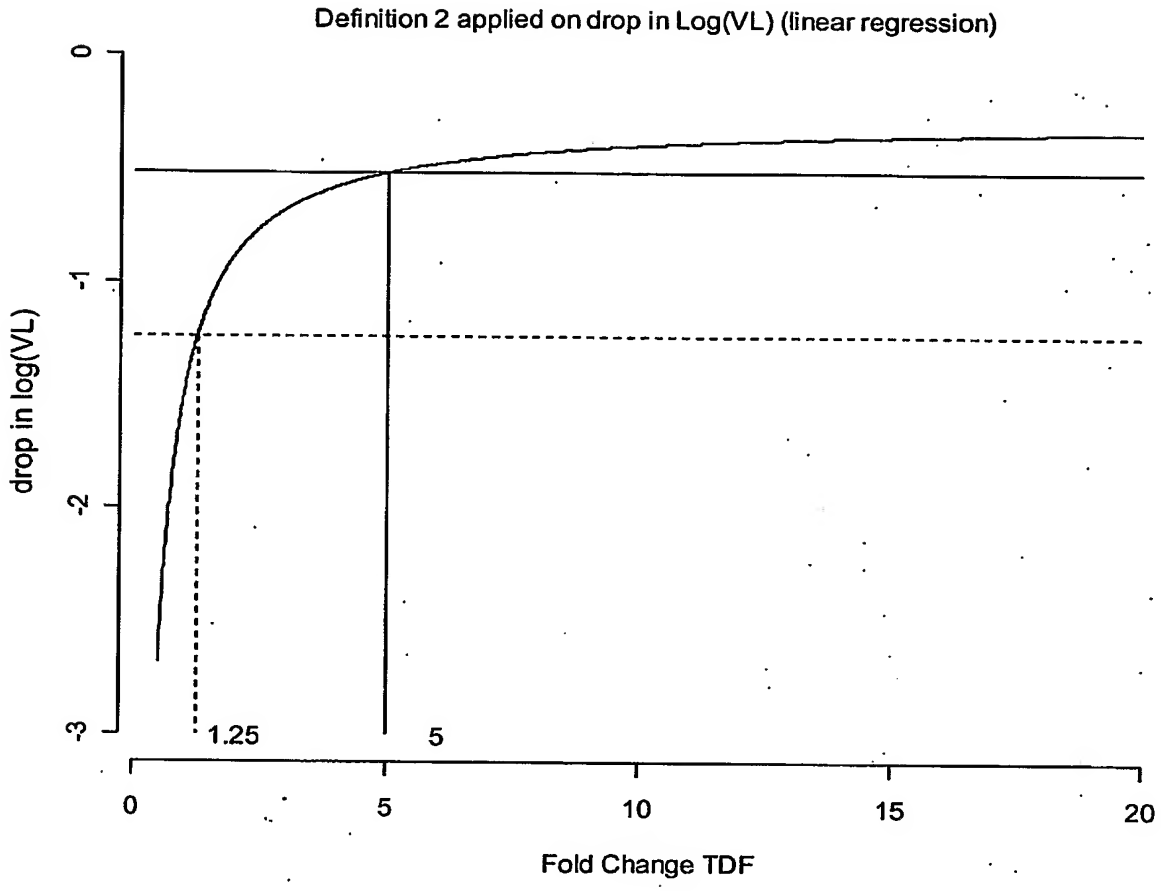

Figure 5

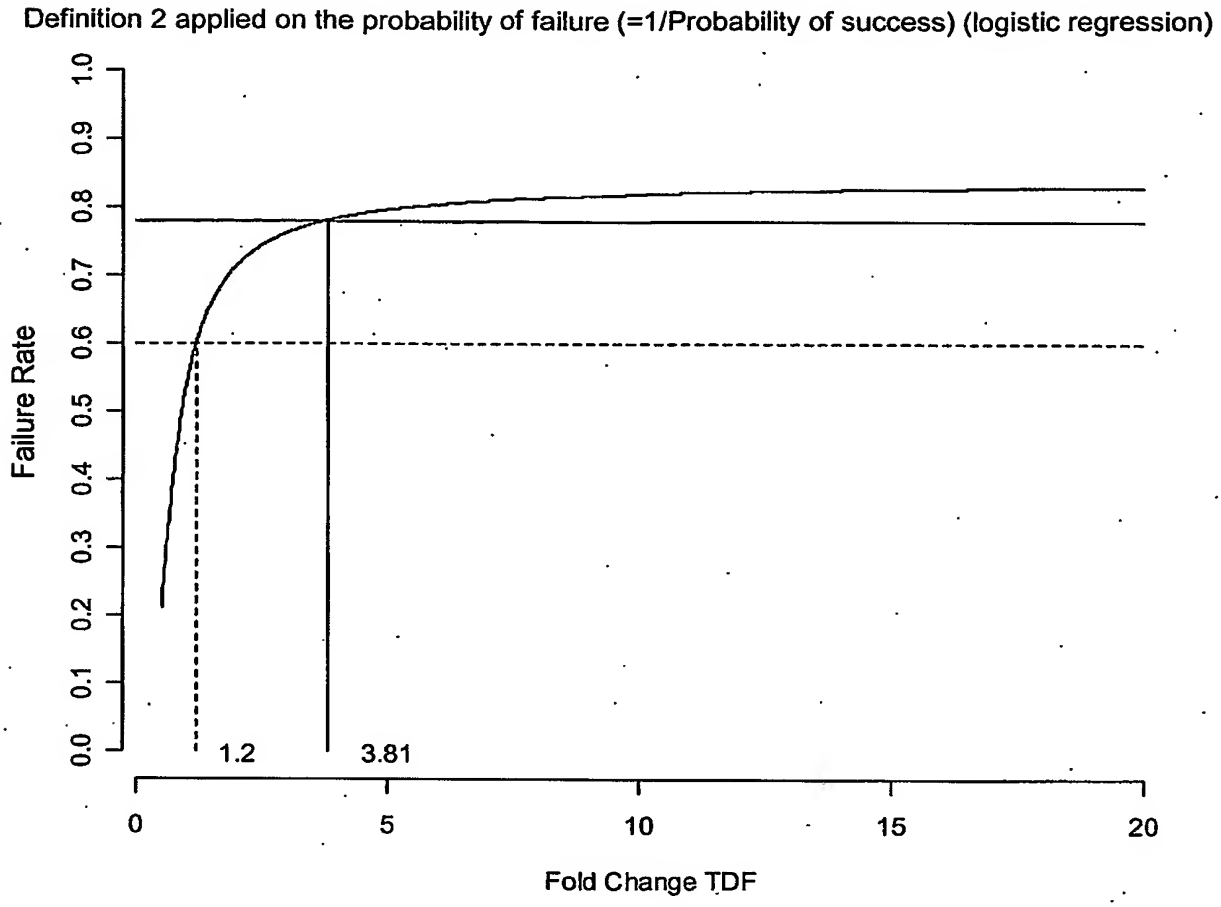

Figure 6

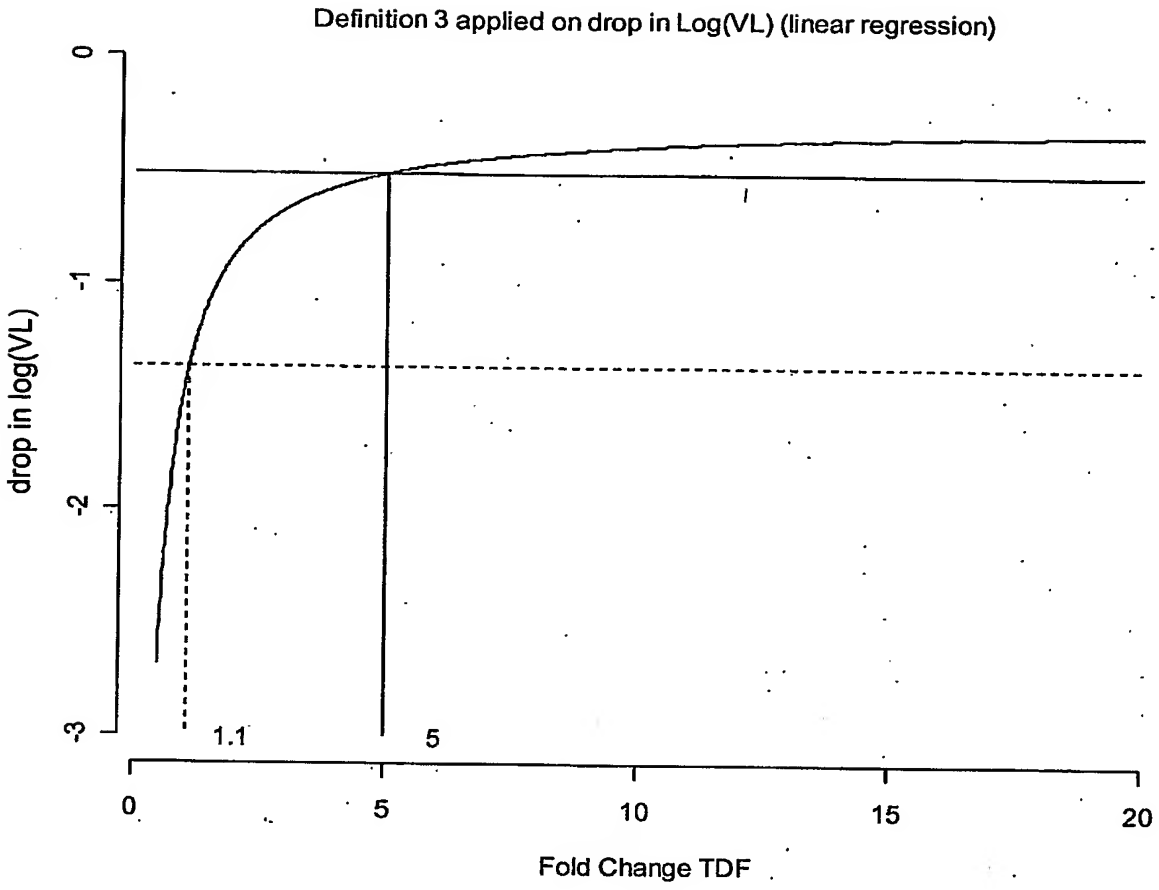

Figure 7

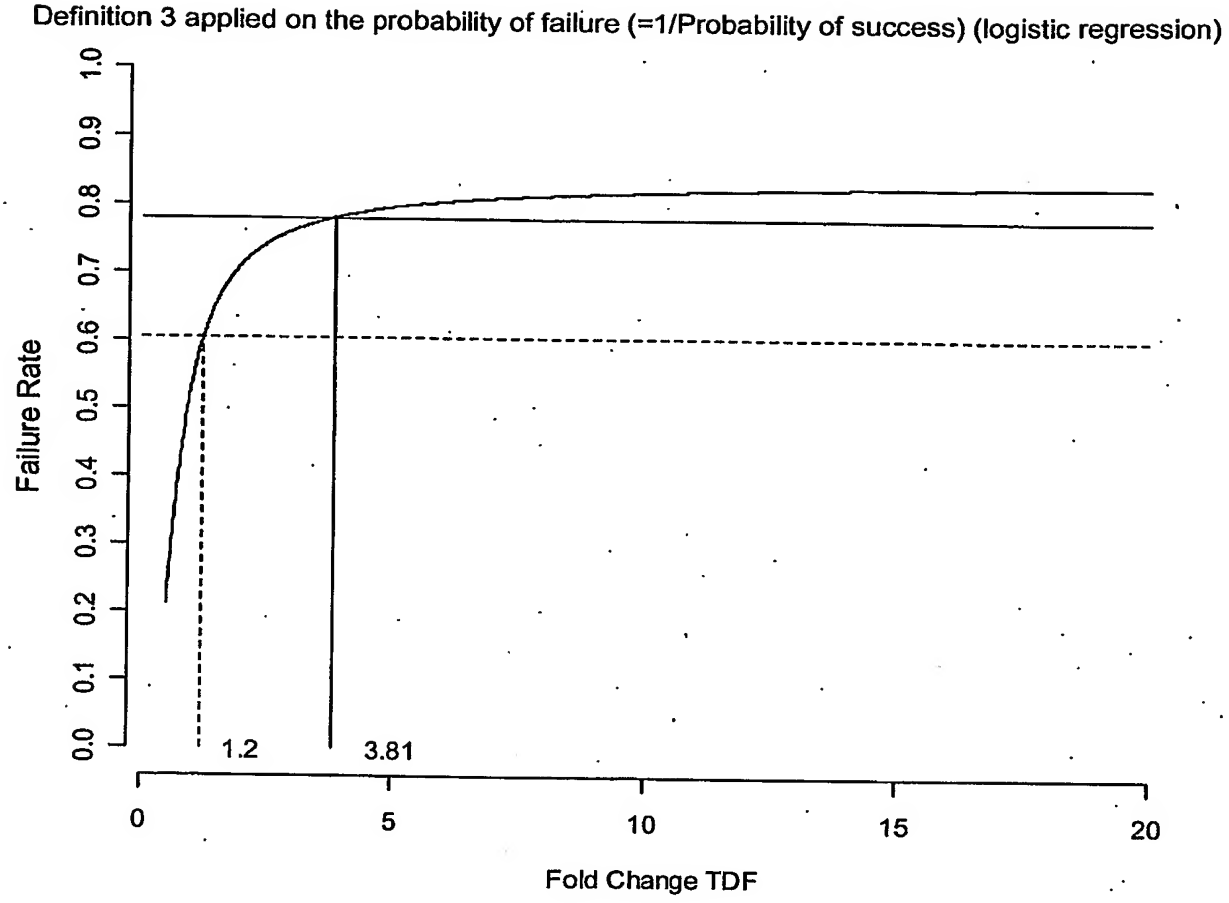

Figure 8